Hardware and software for a PC-based workstation devoted to philological (principally Greek and Latin) studies

Lana, Maurizio

Historical Social Research, Vol. 14 — 1989 — No. 4, 70-75

Hardware and Software for a PC-based Workstation Devoted to Philological (Principally Greek and Latin) Studies

*Maurizio Lana**

1. Near perspectives, but not too near

In this latest years, I used to regard what Professor Bagnall wrote about hardware and software for the classicist (Bagnall 1985) as a relevant reference work, although it is no longer up to date after the technical developments in the fields of CPU, of mass storage, of recording media. According to Bagnall, the ideal - so to say - personal workstation for anyone who wanted to carry out on his own a philological research was Ibycus computer, which was developed by professor T. Brunner from the University of California at Irvine and by D. Packard.

Nowadays, our perspectives have partially changed, although I am quite sure that Ibycus is still an unsurpassed machine in the specific fields where it can best be exploited (that is in the lexical researches on the Greek texts included in TLG - Thesaurus Linguae Graecae - CDROM). Today, in fact, we can build a workstation which has similar (and not exactly alike) potentialities to those of Ibycus but not its intrinsic limit which, 1 dare say, is very severe: Ibycus in fact did not allow us either to transfer any data to other programs nor to process a text or to use a database program. However, Ibycus's speed and efficiency are still unsurpassed.

I have talked of near perspectives because the single elements of the scenery that I describe are already present on the market and in the field of research. The problem, however, concerns which ones to choose and how to combine them and make them work together, which requires some skills; that is why these elements combination has been accomplished in few instances. On the other hand, the userfriendliness of such products cannot but improve with their spreading, as it has already happened for other hardware and software products in the past.

* Address all communications to Maurizio Lana, Str. Del Lauro 47, 1-10132 Torino, Italy.

Historical Social Research, Vol. 14 — 1989 — No. 4, 70-75

2. Matters of Principle

2.1 The machine has to be useful to man, and not vice versa

Someone might wonder if such a statement may start a philosophical rather than a philological discussion. Actually we cannot forget that the flavor of humanism - so to say - which is an aspect of this principle, comes also from the genuine esteem for man that has always enlivened and still enlivens a relevant part of philological research; well, such a principle has been at work in the field of computer science right from its beginnings. I mean that the search for a larger and larger userfriendliness in user's interface of computers (from machine-language towards icon-driven operating systems) must prevail even if it is still diffused the older, contrasting trend, that of being fascinated by the machine, and its magic mechanism and procedures, while attempting to master it.

2.2 A personal computer is often better than a mainframe

A principle that can be, and is debated; but without it, it is much more difficult to achieve the goal of considering the machine a tool at man's disposal. The scholar can choose his ways and times of working much more easily and freely when using a personal computer: he can make up for the longer working out times with the time he spares either by not having to enter a waiting list or to go under time-sharing priorities. Besides, the graphic element for linguistics and philology is at a more advanced stage of development on personal computers rather than on mainframes. On the other hand, the file transfer from personal to mainframe is now part of everyday experience when the required working out powers surpass those of a workstation.

2.3 Do not give up graphic quality

1 can see no point in fatalistically accepting to read - as it often still happens when one works on mainframes - Greek, Sanskrit, Arabian, Hebrew or Syriac transliterated in long sequences of standard ASCII characters and not in their original ones, although one can understand the difficulty of people operating mainframes in meeting the scholar's needs. This habit appears to be still more questionable when we realize that transliteration, often clumsy, makes the use of computer sciences instruments harder and longer as to the learning times for the scholar full of good will but still inexperienced.

2.4 Do not give up philological quality of texts in machine readable form

The texts available on magnetic or optical media are almost always the transcription on a new medium of a previous critical edition in book form, deprived of the apparatus criticus and, as a consequence, of the variants. This implies the loss of a relevant amount of information. That is why the philologist who wants to be such, and to work effectively with new tools, feels it necessary to have at disposal texts in machine readable form including the variants, even if all this gives rise to a new problem: that of defining a standard »de facto«, although not formally defined, to include variants in the text they belong to.

One may wonder, however, whether the amount of information that is lost whenever you cannot include the variants in the text can be neglected when working on huge amount of text from a statistical point of view. Such a problem can be investigated from a close standpoint by means of philological, statistical, mathematical tools if not in a general, exhaustive way, at least as a trial study of representative textual samples. Yet, as far as I know, such attempts have never been made.

2.S. Graphical standardization of texts

Every language has some words that can be written in different ways. Any of these differently written words is, of course, a word »per se« for computer programs. That is not only a lemmatization problem, because in Latin, e.g., it appears principally with unflexed words: think of »quamobrem / quam ob rem«; and so on. But, still speaking of Latin, if scholars studying classical Latin may perhaps accept graphical standardization, those studying medieval Latin accept standardization with great difficulty (or do not accept it at all!) because often different writings convey a great deal of information about the manuscripts' age and place of origin, or about the language's stage of development.

Perhaps producing thesauri containing the different writings of words could help. But how can we know if we arc collecting 'the' different writings or only 'some' different writings? Collecting into a thesaurus 'the' different writings of one word means that one exhaustively know the universe of texts that could possibly contain that word with related writings: a very difficult knowledge!

Another way of facing the problem could be that of giving fuzzy-logic abilities to literary analysis programs. Speaking of 'fuzzy-logic abilities' I really speak about a fuzzy argument, that could be so defined: imitating the human ability to judge similarity between items. Proximity Board and Friendly Finder software (from Proximity Technology Inc. of Fort Lauderdale) are an interesting application of those fuzzy abilities.

One could also think of Artificial Intelligence, expert systems, and so on; but it seems to me that we are still too far from quickly and very effectively usable tools.

3. Description of the Perspectives

3.1. In general, abstract terms

A scholar's workstation should be a personal computer built around a fast and powerful CPU (Intel 80286 or 80386) running MS DOS or OS/2, and should be equipped with:

- RAM memory over 2M (two millions bytes);
- high resolution graphic monitor and display adapter, preferably multistandard;
- high resolution graphic printer
- mass storage device of 30M (thirty millions bytes) at least, which can be sided by a WORM (Write Once, Read Many) drive;
- CDROM drive;
- CDROM textual data base(s)
- text retrieval program which can:
  + display texts graphically, each with his own alphabet:
  + mix different alphabets with their own fonts on the same line;
  + allow queries concerning not only words, but also grammatical categories (e.g. »preposition followed by an infinitive used as a noun«);
  + when required, expand a headword in all or some of his forms;
- word processing program which can:
  + interact with text retrieval program;
  + display every text with his own alphabet (as already said speaking about text retrieval program);
  + combine several diacritical marks on the same alphabetical letter (for example, a Greek vowel with an accent, a breathing, a iota subscript).

3.2 In concrete terms

The essential elements of the scenery 1 have described, those which arc worth dwelling upon, being the ones concerning software. 1 shall not go into details about models and types of CDROM drives, monitors and display adapters, printers and so on.

+ the CDROM textual database suiting the scholar's needs can be either TLG CDROM or CDROM SAMPLER # 1 from PHI (Packard Humanities Institute) and CCAT (Computer Center for Analysis of Texts of Pennsylvania University). TLG CDROM collects Greek texts whereas PHI-CCAT CDROM collects miscellaneous texts from various ancient and modern literatures (among them, there is a great deal of Latin texts).

-I- the text retrieval program can be LBase from Silver Mountain; it is intended to be used in conjunction with texts in the following formats:

- Beta Format (TLG and PHI-CCAT CDROMs)
- BHS format (texts from CCAT)
- morphologically tagged (texts from CCAT)
- parallel aligned masoretic text/LXX (from CCAT)
- ABS Greek New Testament.

LBase also aims at achieving queries based on grammatical categories in the case of texts that have been previously morphologically tagged. On the other hand, Lbase is not able to lemmatize, and lemmatizing, by itself, is not easy at all. As far as Latin is concerned, Doctor Bozzi from the Istituto di Linguistica Computazionale of CNUCE in Pisa has been working about lemmatization with good results within a research project whose director is Professor Marinone from University of Turin. Here at the Conference, Doctor Bozzi speaks about that in his contribution »A Latin Morphological Analyzer«.

+ the word processing program can be Multi-Lingual Scholar from Gamma Productions, a cheap one, too, together with his utility program CTIU (Configurable Text Interchange Utility). This utility allows the transcoding from any sequence of one or more characters to any other sequence of one or more characters. Although it may be configured by user, CTIU is delivered ready to operate linking any two among the following formats:

- ASCII
- WordStar
- NotaBene
- LBase
- WordCruncher
- standard transliteration of Hebrew, Russian, Arabic, Syriac, and so on.

A very interesting program, mixing text retrieval and word processing, and being able to directly access a CDROM drive, is Pennwrite, from Pennsylvania University; but I received it too late to be able to use it and so to speak about it adequately. There are other scholarly word processing programs, and the first among them is NotaBene from Dragonfly Software, but the most relevant and unique feature of TLG CDROM, Lbase and Multi-Lingual Scholar is that they are studied to interact and work

together. My quoting them in this contribution has no advertising aim, then.

4. Conclusion. Which modifications will the tools here described produce in scholarly research?

One can wonder which space is left by the present software cornucopia to scholars: perhaps, the old and worthy activity of browsing trough texts to create repertories, concordances, indexes, in one word: creating tools for other scholars, can no more by itself point out the clever scholar's skills.

Today, perhaps, they must show themselves in thinking of (and also solving!) subtler and more complex questions and hypotheses, needing data that can be collected only with specific literary, text retrieval software.

References

BAGNALL, R. 1985. »Word Processing for the Classicists APA Committee on Education. Educational Papers 2